Jacek Lech

Moda na średnią medianę

Podobno statystyk, trzymając głowę w lodówce i nogi w rozgrzanym piekarniku, twierdzi, że średnia temperatura jest całkiem znośna.

Wśród pojęć statystycznych najczęściej słyszymy o „średniej", przez którą potocznie rozumie się średnią arytmetyczną. W szkołach liczymy średnią ocen, w sprawozdaniach sportowych spotykamy się ze „średnią skutecznością" koszykarzy czy piłkarzy, a statystycy przerażają nas średnim spożyciem alkoholu przez jednego Polaka.

Zrozumienie tego pojęcia zwykle nie przysparza uczniom kłopotów. Jeśli podczas pewnego turnieju piłkarskiego rozegrano 11 meczów i zdobyto w nich następujące liczby bramek: 1, 4, 2, 2, 8, 0, 2, 4, 4, 4, 1, to średnia liczba bramek strzelonych w jednym meczu wynosi:

$$\frac{1+4+2+2+8+0+2+4+4+4+1}{11} \approx 2{,}9.$$

Czytając w relacji z tych rozgrywek, że średnio zdobyto „blisko 3 bramki" w jednym spotkaniu, mamy prawo uznać, że mecze były atrakcyjne.

Jednak bywa, że średnia nie opisuje dobrze zestawu danych, a nawet zdarza się, że wprowadza w błąd. W starym dowcipie pewien statystyk utonął w jeziorze o średniej głębokości 1 metra.

Moda, czyli wartość typowa

Wyobraźmy sobie, że właściciel firmy krawieckiej chce się dowiedzieć, z iloma kieszeniami ma szyć spodnie, by cieszyły się jak największą popularnością. W tym celu spytał pewną liczbę ludzi, ile kieszeni powinny mieć ich wymarzone spodnie. Aby jego sonda miała jakąś wartość, powinien zebrać odpowiednio dużą liczbę danych, ale dla uproszczenia załóżmy, że poprzestał na jedenastu odpowiedziach: 1, 4, 2, 2, 8, 0, 2, 4, 4, 4, 1.

Czy średnia (2,9 kieszeni) jest wartością, którą powinien kierować się właściciel firmy? Raczej wątpliwe. Trudno uszyć spodnie z 2,9 kieszeniami, a jeśli zdecyduje się na 3 kieszenie, to może się zdarzyć, że nie znajdzie na nie nabywcy! Przecież nikt z pytanych nie marzył o trzech kieszeniach. Krawiec zrobiłby najrozsądniej, wybierając najczęściej spotykaną wśród odpowiedzi liczbę kieszeni. Przy większej liczbie danych należałoby je najpierw uporządkować według wielkości lub pogrupować w klasy, ale w tym przypadku łatwo zauważyć, że ową najpopularniejszą liczbą kieszeni jest 4. Sonda wskazuje, że spodnie z czterema kieszeniami cieszyłyby się największym wzięciem. Ta wielkość, która występuje wśród danych największą liczbę razy, nazywa się „modą" albo „dominantą".

Może się zdarzyć, że na liście danych znajdzie się nie jedna, a kilka danych o takich samych, największych częstościach występowania. Wtedy każdą z nich nazywa się modą. Zestaw danych może mieć zatem kilka dominant.

Mediana

Inną sytuacją, w której średnia – zamiast informować – dezinformuje i (zwłaszcza nas, nauczycieli) irytuje, są raporty o średnich zarobkach. Łatwo obliczyć średnią płacę w małej firmie zatrudniającej cztery osoby: trzech pracowników zarabiających po 1000 złotych i szefa zarabiającego 5000 złotych miesięcznie. Wynosi ona 2000 złotych, chociaż 75% zatrudnionych otrzymuje zaledwie połowę tej sumy. Zdecydowanie w tym przypadku informację o średniej należy uzupełnić wartością mody równej 1000 złotych. Jednak nawet średnia w towarzystwie mody może wprowadzić w błąd.

Zarobki pracowników firmy „Coto” są następujące: 500, 500, 600, 800, 800, 900, 1000, 1000, 1000, 1100, 1500, 1600, 5000, 5000, 5000, 5000, 5000, 8000, 10 000 (złotych). Szukając pracowników, dyrektor może ogłaszać, że średnia płaca w jego firmie wynosi około 2860 złotych, a najczęstszą płacą (modą) jest 5000 złotych. Jednak łatwo zauważyć, że te informacje nie oddają pełnego obrazu listy płac firmy.

Jeszcze więcej może ukryć średnia tak olbrzymiej liczby danych, jak np. zarobki strefy budżetowej.

W takich sytuacjach wielkością niosącą ważną informację jest mediana. „Mediana” znaczy po łacinie „środkowa” i rzeczywiście jest to środkowa wartość na liście danych uporządkowanych według wielkości. Jeśli liczba danych jest parzysta, to medianą jest średnia z dwóch danych najbliższych środka. Można stwierdzić, że 50% danych jest nie większe niż mediana.

n – nieparzysta liczba danych
0, 1, 1, 2, 2, 2, 4, 4, 4, 4, 8
mediana
medianą jest środkowa,
a więc wielkość o numerze (n+1)/2
na uporządkowanej liście danych

n – parzysta liczba danych
1, 1, 2, 2, 2, 4, 4, 4, 4, 8
„środkowe” wartości
mediana= =(2+4):2=3
mediana jest średnią arytmetyczną,
wielkości o numerach n/2 i n/2+1
na uporządkowanej liście danych

Wcześniej przedstawiona lista płac firmy „Coto” jest uporządkowanym zestawem 19 liczb, zatem środkową jest dziesiąta wartość. Wynika stąd, że medianą jest 1100 złotych, a więc 50% pracowników firmy nie zarabia więcej niż 1100 złotych. Warto zwrócić uwagę, jak bardzo różne wartości mają średnia, moda i mediana dla tej listy płac i jak niedoskonały obraz rzeczywistej sytuacji daje znajomość tylko jednej z tych wielkości. Może więc powinniśmy nalegać, aby wśród informacji na temat zarobków w „budżetówce” podawać, oprócz średniej, również modę i medianę? Tylko kto zrozumie znaczenie tych wielkości?
I dlatego warto uczyć statystyki. ■